

were taken after each increment of 0.01 or 0.02 ml to a solution of 10 or 15 ml, which was stirred constantly with a Teflon-coated magnetic bar. Below 15° about 1–2 min were required to reach a constant pH reading after each addition of the NaOH solution.

Sedimentation. All sedimentation experiments were performed with a Beckman-Spinco Model E analytical ultracentrifuge. Sedimentation velocity runs were done in single sector cells with Kel F centerpiece and quartz windows and with Schlieren optics. For low-speed sedimentation equilibrium runs, Rayleigh optics were used and the solution and solvent were filled in double sector cells with Kel F centerpieces and sapphire windows. Double-sector capillary centerpieces were used in synthetic boundary measurements. Sedimentation plates were analyzed with a Nikon micro-comparator. The number and positions of the fringes for the sedimentation equilibrium runs were read and used to calculate the

apparent weight- and Z-average molecular weights, M_w^{app} and M_z^{app} , from a computer program by Dr. D. C. Teller.³⁴ The reciprocals of M_w and M_z were determined by extrapolating to zero concentration the straight lines of $1/M_w^{app}$ vs. the mean concentration (the average of the concentrations at the meniscus and bottom) and of $1/M_z^{app}$ vs. twice the mean concentration.

Acknowledgment. We thank Professor L. Peller for several valuable discussions and Professor H. Martinez for providing the computer program for the solution of the Zimm–Bragg equation.

(34) D. C. Teller, Doctoral Dissertation, University of California, Berkeley, Calif., 1965.

A Computer Investigation into the Origin of the Code

Marc S. Rendell

Contribution from the Department of Medicine and Program in Biophysics, State University of New York Downstate Medical Center, Brooklyn, New York 11203. Received September 28, 1971

Abstract: Quantum mechanical methods are used to simulate interactions between glycine and polynucleotide base sequences. High-magnitude, configurationally specific interactions with features similar to Watson–Crick base pairing are obtained. Glycine manifests preferential affinities in complexing with the stacked base pairs. The feasibility of selective interaction between polar amino acids and polynucleotide base sequences is considered. The implications of such interactions for the origin of the genetic code are discussed.

1. Introduction

The genetic code has two functions. The first is to preserve information existing in polynucleotide sequences by strict base complementarity upon replication. The other is to permit usage of available information through the relation of a specific amino acid to a given trinucleotide sequence during translation.

The forces subserving the first function are those of Watson–Crick base pairing and are well understood. The specificity of the physical association of adenine to thymine and uracil and of guanine to cytosine has been confirmed both theoretically^{1–3} and experimentally.^{4–6} Specificity in this case has been demonstrated to be due to the varying feasibility of hydrogen bonding between the charged groups of the bases.

Unlike base pair complementarity, the physical basis for the translation of codon into amino acid remains unclarified. It is clear that distinctive, selective physical interactions between polynucleotide bases and amino acids, either singly or as the side chains of polypeptides, must be involved. In the contemporary translation mechanism, these physical interactions must take place in a complex system consisting of a given tRNA and its complementary aminoacyl synthetase. The crucial

interactions involved here most likely reflect the high degree of structure in this system.

The process of translation in the primitive coding system, unlike the contemporary mechanism, could not have taken place in a highly structured environment. Indeed, the primordial coding system arose as a consequence of physical interactions occurring among the relatively unstructured polymers and polymeric subunits in the primitive “soup.”^{7–10} Here, base pairing ensured conservation of informational sequences. There is far less certainty as to the type of interactions responsible for the primitive translational process. Experimental evidence on this point has only recently begun to appear with the demonstration that polyarginine and polylysine exhibit preferential binding to polynucleotides of varying base composition.^{11,12}

In this regard, theoretical investigation indicates that selective interactions between glycine and the individual nucleotide bases are quite feasible and are due to selective component interactions between the carboxyl and amino groups of glycine and oppositely charged moieties of the bases.^{13,14} In this paper, it is shown that glycine is capable of selective, configurationally

(1) B. Pullman, P. Claverie, and J. Caillet, *Proc. Nat. Acad. Sci. U. S.*, **55**, 904 (1966).

(2) R. Rein, P. Claverie, and M. Pollak, *Int. J. Quantum Chem.*, **2**, 129 (1968).

(3) H. A. Nash and D. F. Bradley, *Biopolymers*, **3**, 261 (1965).

(4) J. S. Binford, Jr., and D. M. Holloway, *J. Mol. Biol.*, **31**, 91 (1968).

(5) R. A. Newmark and C. R. Cantor, *J. Amer. Chem. Soc.*, **90**, 5010 (1968).

(6) Y. Kyogoku, R. C. Lord, and A. Rich, *ibid.*, **89**, 497 (1967).

(7) F. H. C. Crick, *J. Mol. Biol.*, **38**, 367 (1968).

(8) L. E. Orgel, *ibid.*, **38**, 381 (1968).

(9) C. R. Woese, *ibid.*, **43**, 235 (1969).

(10) A. M. Lesk, *J. Theor. Biol.*, **27**, 171 (1970).

(11) J. C. Lacey and K. M. Pruitt, *Nature (London)*, **223**, 799 (1969).

(12) S. W. Fox, J. C. Lacey, and A. Yuki at the Surface and Colloid Chemistry Symposium, Toronto, Canada, 1970.

(13) R. Rein, M. S. Rendell, and J. P. Harlos in “Molecular Orbital Studies in Chemical Pharmacology,” by L. B. Kier, Ed., Springer Verlag, New York, N. Y., 1971.

(14) M. S. Rendell, J. P. Harlos, and R. Rein, *Biopolymers*, **10**, 2083 (1971).

specific interactions with polynucleotide base sequences. The implications of such interactions for the origin of a primitive mode of translation are drawn.

2. Method

The total interaction between glycine and each nucleotide base is computed as a sum of monopole-monopole energies (electrostatic interaction) and bond polarization potentials due to monopole-bond polarization interaction) and bond-bond polarization (dispersion interaction). The monopole charge distributions (net atomic charges) were calculated previously using the IEHT SCF-MO method.^{13,14} The equations used for the computation of the interaction components were as originally developed in ref 2: (a) the electrostatic energy given by the sum of the Coulombic matrix components for the atomic partial charges

$$E_{\text{elstat}} = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} Q_i Q_j / R_{ij}$$

(b) the polarization energy resulting from the net atomic point charges of molecule 1 perturbing the bonds of molecule 2²

$$E_{\text{pol}}^{1 \rightarrow 2} = -1/2 \sum_{k=1}^{b_2} p_k^T (\bar{F}_k \cdot \bar{F}_k) + d_k (\bar{F}_k \cdot \bar{p}_k^L)^2$$

$$\bar{F}_k = \sum_{i=1}^{n_1} (Q_i / R_{ik}^3) \bar{R}_{ik}$$

where the summation is over the number of bonds in molecule 2, b_2 , and $d_k = p_k^L - p_k^T$ where p_k^L and p_k^T are the longitudinal and transverse polarizabilities of the k th bond. \bar{R}_{ik} is the vector connecting atom i of molecule 1 with the midpoint of bond k , and R_{ik} is its magnitude. p_k^L is a unit vector pointing along bond k . The energy of polarization of bonds of molecule 1 by the monopoles of molecule 2 is given by a symmetric expression. (c) The dispersion energy is given by²

$$E_{\text{disp}} = -1/4 I_1 I_2 / (I_1 + I_2) \sum_{j=1}^{b_2} \sum_{i=j}^{b_1} 1/r_{ij}^6 [6 p_i^T p_j^T + p_i^T d_j (3(\bar{p}_j^L \cdot \bar{r}_{ij})^2 + 1) + p_j^T d_i (3(\bar{p}_i^L \cdot \bar{r}_{ij})^2 + 1) + d_i d_j (3(\bar{p}_i^L \cdot \bar{r}_{ij})(\bar{p}_j^L \cdot \bar{r}_{ij}) - (\bar{p}_i^L \cdot \bar{p}_j^L)^2)]$$

where \bar{r}_{ij} is a vector directed from the midpoint of bond i to the midpoint of bond j and I_1 and I_2 are the molecular potentials obtained from the SCF-MO of the molecules.

The interaction is carried out by scanning configurational space as defined by glycine and the nucleotide bases for minimum potential energy. The molecular dimensions of glycine are such that its interactions with a polynucleotide base sequence may be viewed as discrete approaches to stacked base pairs; in other words, it fits the 3.4 Å space between the bases of a stacked pair. The coordinates of the base pairs with respect to the helix axis were given by crystallography,¹⁵ the position of the upper base being given by a 36° counterclockwise rotation with respect to the helix axis and a z axis translation of 3.4 Å. Scanning space consists of the three coordinate axes and the three interaxis rotations: $\theta, x \rightarrow y; \gamma, y \rightarrow z; \phi, z \rightarrow x$. The glycine molecule is first rotated in units of 5° and then translated

(15) R. Landridge, D. A. Marvin, W. E. Seeds, H. R. Wilson, C. W. Hooper, M. H. F. Wilkins, and L. D. Hamilton, *J. Mol. Biol.*, **2**, 38 (1960).

along the axes in units of 0.2 Å. The closest approach permitted between two monopoles is set at 2 Å. At this distance shell overlap may be neglected. The scan is carried out for the electrostatic component in this way. When an electrostatic minimum has been obtained, a reduced scan with units of 2° and 0.1 Å is performed to give the best configuration for the total energy.

3. Results

For each of the 16 base pairs, a well-defined configuration of maximum interaction as evidenced by a deep potential well is obtained. Consideration of the interaction matrix elements reveals that the interactions are principally due to the large negative charge on the carboxyl group and the large positive charge on the amino group complexing with the oppositely charged keto oxygens and ring nitrogens and the positively charged amino hydrogens of the bases. This induces a high degree of directionality of complexing which, of course, is a factor in producing selectivity. The energies are given in Table I, each stacked pair being

Table I. Energy Components for Stacked Base Pairs (kcal/mol), Given as Positive Interaction

Base pair	Figure	E_{elstat}	E_{disp}	E_{pol}	Total
GC	1(t)	45.5	22.2	25.2	92.9
GG	3(b)	51.3	14.2	19.4	84.9
CC	3(t)	45.2	16.0	21.7	82.9
CA		35.1	19.9	23.8	78.8
AC		34.9	21.4	21.0	77.3
UG		39.3	11.7	15.9	66.9
GU		38.3	9.4	13.5	61.7
AA		27.6	12.9	15.8	56.3
CG	1(b)	29.0	9.8	14.6	53.4
CU	4(b)	25.5	10.9	12.3	48.7
UA	2(t)	20.8	14.1	13.1	48.0
UC	4(t)	25.5	8.9	12.0	46.4
GA	5(t)	25.3	8.7	10.1	44.1
UU		23.6	8.4	10.7	42.7
AG	5(b)	21.6	7.8	10.3	39.7
AU	2(b)	19.3	6.5	9.5	35.3

labeled as the bottom base followed by the upper one. The magnitudes decrease from 93 kcal for the interaction with GC (Figure 1) to 35 kcal for that with AU (Figure 2), the difference being almost 60 kcal. The base pairs CC (Figure 3), CA, and AC may be considered to give interactions of high magnitude as does GC while CU, UC, UA, GA, AG, UU, and AU give interactions of low magnitude. The energy difference between the area of high magnitude and that of low magnitude is almost 40 kcal.

In addition to the specificity evidenced by these large energy differences, the directionality of complexing is very important in ensuring selectivity. For example, whereas GC gives the largest interaction of 93 kcal, CG (Figure 1), with a rotated relative disposition of the same pertinent charged groups relative to GC, gives an interaction of only 53 kcal. Similarly, the interaction with UA is 48 kcal while that with AU is 35 kcal (Figure 2). Another quite remarkable example of the importance of directionality in complexing is provided by the configurations of maximum interaction with the pairs

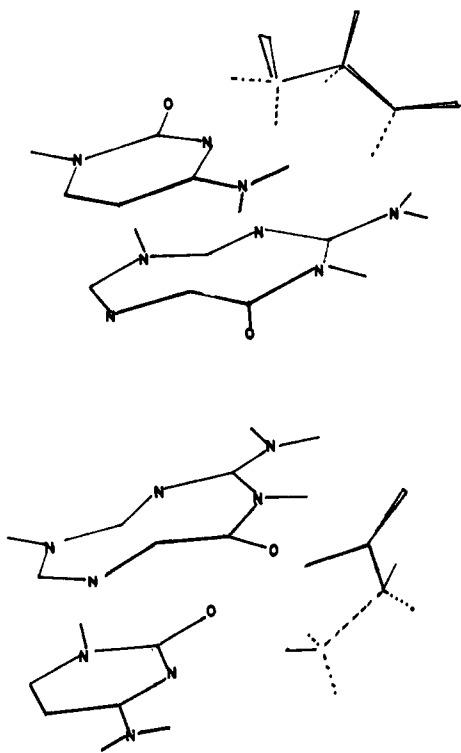


Figure 1. Top, glycine interaction with guanine-cytosine in helix space (double bonds and purine fusion bond have been omitted from sketch). Bottom, interaction with cytosine-guanine.

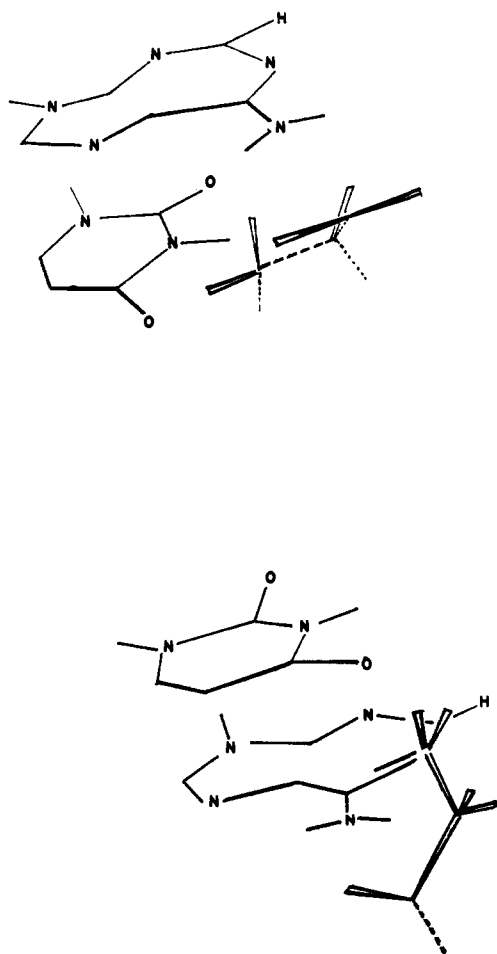


Figure 2. Top, UA. Bottom, AU.

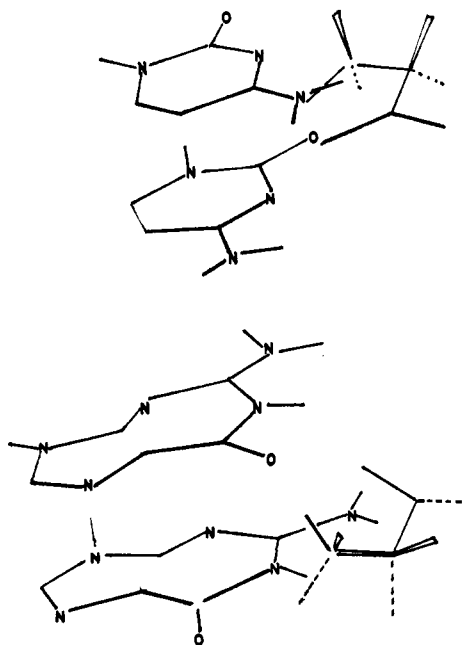


Figure 3. Top, interaction with cytosine-cytosine. Bottom, interaction with GG.

CU, UC, GA, and AG. In each case, the configuration of maximum interaction assumed is basically a complex with only one of the bases (Figures 4 and 5). The matrix components for the other base are actually repulsive (Table II). Glycine tends to assume a position as far away as possible from uracil or adenine while orienting for the maximum interaction with cytosine or guanine. This is quite significant in that gly-

Table II. Individual Base Contributions for Certain Pairs^a

Base pair	Base	E_{elstat}	E_{disp}	E_{pol}	Total
GC	G	22.5	18.2	17.5	58.2
	C	23.0	4.0	7.7	34.7
CG	C	14.6	3.5	5.7	23.8
	G	14.4	6.3	8.9	29.6
CU	C	25.2	10.4	11.8	47.4
	U	0.3	0.5	0.5	1.3
UC	U	-0.3	0.7	1.3	1.7
	C	25.8	8.2	10.7	44.7
UA	U	10.8	3.4	5.1	19.3
	A	10.0	10.7	8.0	28.7
AU	A	3.4	3.0	3.8	10.2
	U	15.9	3.5	5.7	25.1
GA	G	26.7	8.5	9.8	45.0
	A	-1.4	0.2	0.3	-0.9
AU	A	-4.1	1.1	1.6	-1.4
	G	25.7	6.7	8.7	41.4
GU	G	27.0	6.4	9.2	42.6
	U	11.8	3.0	4.3	19.1
UG	U	13.3	2.9	4.5	20.7
	G	26.0	8.8	11.4	46.2

^a Given as positive interaction. Note contribution of uracil to UC and CU and of adenine to AG and GA.

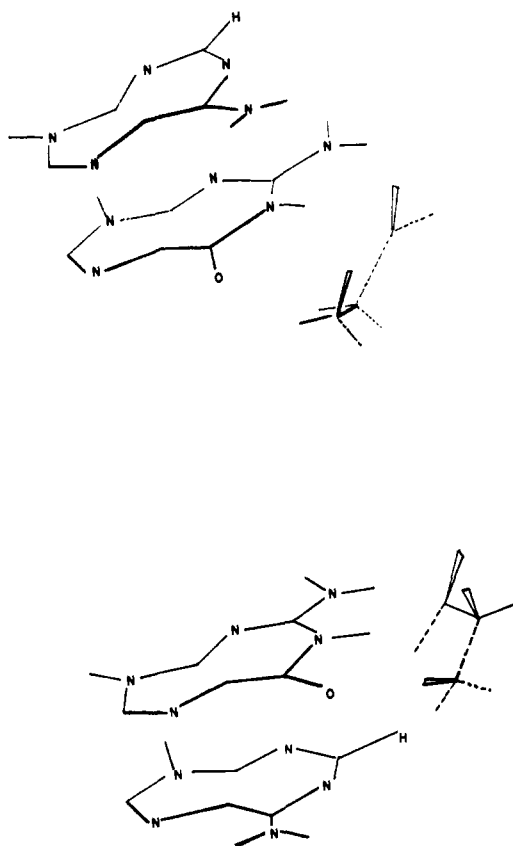


Figure 4. Top, interaction with AG. Bottom, GA. Note that, in maximizing interaction, glycine has assumed a position as far from adenine as possible and is, indeed, in the guanine plane.

ine would tend not to interact with these base pairs on a polynucleotide base sequence since maximization of interaction would lead it into the space of the neighboring base pairs. For example, given a sequence of UCC on a polynucleotide, interaction with UC would lead glycine into the space of CC and encourage complexing with it. However, a sequence of UCU could lead to a complex essentially with cytosine alone.

4. Discussion

Calculations have been performed which, in essence, parallel previous theoretical studies by various investigators on Watson-Crick base pairing.^{1,2} Just as in these previous studies, selective interactions due to the complexing of charged groups are obtained for glycine and the base pairs.

Several aspects of the present calculations deserve emphasis. The first is that the magnitude of the interactions computed is quite large. By way of comparison, theoretical studies on Watson-Crick pairing give energies of about 7 kcal for the A-T interaction and 19 kcal for the G-C interaction.^{1,2} A quite direct illustrative comparison may be made, for example, between the glycine complex with CC and the complex with CC on one nucleic acid strand and GG on its complementary strand of the double helix. The energy of interaction for the latter is given by the sum of the two horizontal G-C interactions plus a stacking component of about 8 kcal,² thus a total of about 45 kcal. In contrast, the CC complex with glycine gives an interaction almost twice as large: 83 kcal.

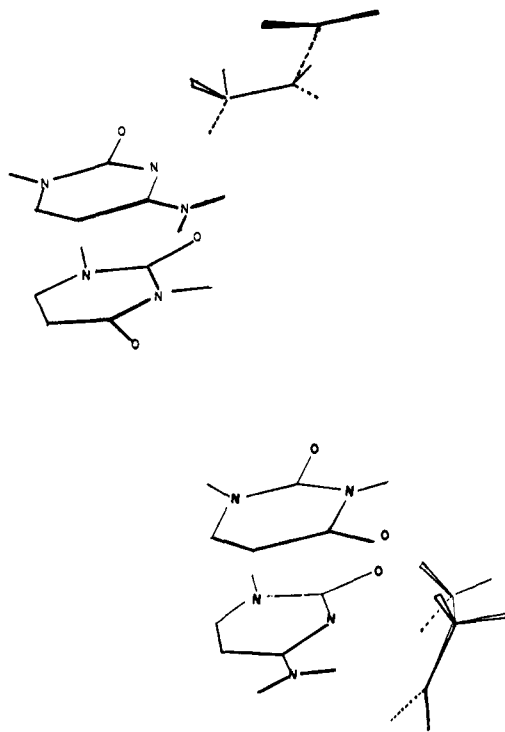


Figure 5. Top, UC. Bottom, CU. Note glycine in cytosine plane at bottom and above cytosine plane at top.

Another factor that should be given equal weight with the large magnitude of interaction is the large difference in magnitude that occurs among the base pairs. The dropoff of 40 kcal from the pairs giving high-magnitude interactions to those giving smaller interactions is of obvious significance in anticipating selective complexing.

Just as important in promoting selectivity as the differences in interaction magnitude is the configurational specificity exhibited. This high directionality in complexing is due, as is that in Watson-Crick pairing, to the fact that the highly charged groups dominate the interaction.

The large energies, the differences in magnitude, and the configurational specificity, taken together, lead to the conclusion that there was in the primordial system the inherent potential for selectivity in the complexing of glycine and polynucleotide sequences. For example, one of the predictions that may be drawn as to the relative rates of complexing is that the order poly(G) \geq poly(C) > poly(A) > poly(U) would obtain. Another is that the rate of complexing to polynucleotides with a high proportion of high-energy pairs such as GC and CC would be greater than to polynucleotides with a preponderance of pairs giving smaller interactions such as UU and UA, and, in particular, pairs giving essentially no interaction such as UC and GA.

It is incumbent for the purpose of experimental verification to consider the anticipated effect of solvent on these interactions. Here, once again, comparisons with Watson-Crick pairing are helpful. Predictions of preferential affinity based on calculations performed on Watson-Crick pairs *in vacuo*¹⁻³ (dielectric constant of 1) have been found to carry over nicely in nmr and ir studies in CHCl_3 and DMSO.^{4-6,16,17} The interactions

(16) R. M. Hamlin, Jr., R. C. Lord, and A. Rich, *Science*, **148**, 1734 (1965).

computed for glycine are similar to those of base pairing in their domination by similar polar groups and should therefore be demonstrable in like manner.

It is natural to inquire what predictions may be made as to the interaction of amino acids other than glycine with nucleotide base sequences. Glycine is the simplest amino acid in that it lacks a side chain. It was chosen for this study due to its polar character and since a side chain would have introduced side-chain bond rotations which would have increased the degrees of freedom in the configurational scan to impracticable levels. Still without attempting expansive scans, it is possible to use glycine interactions as a natural model for the anticipated interactions of amino acids with a polar side chain such as lysine, glutamic acid, ornithine, aspartic acid, and arginine. This is valid since it has been shown that the highly charged groups such as amino and carboxyl dominate the interactions. In essence, a polar amino acid such as lysine may be considered to interact as glycine plus another charged group at a molecular position determined by the conformation of the side chain, most likely variable. The molecular dimensions of such an amino acid are such that it would be expected to interact with a stack of three bases rather than two as for glycine. That the interactions of an amino acid such as lysine or glutamic acid would be different from those of glycine is intuitively apparent as the interaction of three point charges and a given field is different from that of just two of those charges and the same field. As an illustration, the interaction of lysine with a stack XYZ can be viewed as the interaction of glycine with XY plus the interaction of an amino group with base Z. For example, take XY to be AA. For glycine, the electrostatic component is about 30 kcal. The electrostatic interaction of the amino group with base Z will add to this as much as 30 kcal if Z is C or A, 25 kcal if it is U, and 15 kcal for G. Whatever the eventual configurations of maximum interaction arrived at for lysine, those same configurations would evidently not be the best for glutamic acid with an oppositely charged group on its side chain. These considerations serve to illustrate that the presence of a polar side chain is likely to create a whole new set of affinities, and this implies selectivity in the complexing of polar amino acids and nucleotide base sequences just as has been suggested for glycine.

While it is possible to use glycine as a model for the interactions of the polar amino acids this cannot be done when it comes to amino acids with nonpolar side chains such as alanine, valine, leucine, and phenyl-

alanine. Aside from the fact that the presence of a side chain gives rise to steric blocking of many of the complexing configurations found for glycine, nothing should really be said on the basis of the present study as to possible interactions of nonpolar amino acids. The reason for this is that whereas the interactions of polar amino acids may be viewed as similar to the hydrogen bonding of Watson-Crick pairing, the features of which are well preserved in solvent, the effect of an aliphatic side-chain group in modifying the solvation of the amino and carboxyl groups of the glycine moiety is probably crucial to the interaction. In particular, while hydration spheres are "well ordered" around polar groups, this would not be the case around a methyl group. The use of the techniques of this paper for aliphatic amino acids would not do justice to the question of different solvation of these amino acids. However, if future developments do make the electronic structure of hydration in the presence of aliphatic groups accessible, it is not unlikely that differential affinities in complexing with polynucleotide base sequences would also be computed and that these affinities would probably be quite different from those of glycine.

In conclusion, the results of this paper suggest that there exists the potential for selective complexing between glycine and polynucleotide base sequences. In addition, it is presumable that selectivity may likewise prevail in the interactions of several other polar amino acids and polynucleotide sequences. On the basis of such complexing it is reasonable to suppose that a crude selective process of polypeptide templating could arise on polynucleotide sequences. Naturally, such a crude templating process could not be expected to exhibit the efficient, error-free character of present-day mechanisms; in fact, degeneracy may well have been present in primitive coding. Nonetheless, the effect of the postulated interactions would have been to bring about an association of selective character between the processes of nucleic acid replication and polypeptide formation in the primitive system and to thereby give direction to processes of otherwise random polymerization. This would have been a first step in the evolution of the code as it is today.

Acknowledgment. The author wishes to thank Robert Rein for his encouragement in the pursuit of this work. A generous amount of computer time was extended by William Siler and the Department of Computer Science of the Downstate Medical Center under NIH Grant No. RR-00291. In addition, William Siler and Paul Dreizen are thanked for helpful suggestions.

(17) K. Katz, and S. Penman, *J. Mol. Biol.*, **15**, 220 (1966).